

UNITED STATES DISTRICT COURT
DISTRICT OF MASSACHUSETTS

FILED UNDER SEAL

UNITED STATES OF AMERICA)	
)	
v.)	CRIMINAL NO. 13-10200-GAO
)	
DZHOKHAR TSARNAEV)	

**REPLY TO THE GOVERNMENT’S OPPOSITION TO
MOTION TO EXCLUDE TOOLMARK IDENTIFICATION EVIDENCE AND
REQUEST FOR DAUBERT HEARING**

Examination of the grounds the government posits as support for the reliability of toolmarks testimony reveals each to be ephemeral. When stripped to its essence, toolmarks opinion evidence does not satisfy any of *Daubert*’s reliability factors, is inherently unreliable, and as a consequence cannot be admitted.

A. TESTABILITY

The government argues that toolmarks examination is a tested methodology for *Daubert* purposes because laboratory protocols require verification by a second examiner and “numerous studies have been conducted to test the validity and reproducibility of toolmark identification.” Gov’t Opp, at 6-8 (*citing United States v. Otero*, 849 F. Supp. 2d 425, 432 (D.N.J. 2012)). Close scrutiny of the government’s claims, however, shows that there is little merit to either assertion. The *Daubert* Court made clear that the central concern at the threshold is the expert opinion’s “evidentiary reliability.” The Court explained that in the expert context, evidentiary reliability equates to “scientific reliability.” *Daubert*, 509 U.S. at 590 n.9. Science draws a sharp distinction between

reliability and validity, a and this distinction is of great importance to the issue of whether firearm comparison evidence is admissible under *Daubert* and Rule 702. Reliability refers to the extent to which a measuring instrument produces the same result when it is used repeatedly to measure the same object or event. See John I. Thornton & Joseph L. Peterson, *The General Assumptions and Rational of Forensic Identification*, in *Science in the Law: Forensic Science Issues* 19 (David L. Faigman et al. eds. 2012); *United States v. Horn*, 155 F.Supp.2d 530, 538-539 (D. Md. 2002). Consequently,

a parade of forensic scientists who make the same subjective judgment, or a series of machines that give the same readings in response to the same evidence sample, *can only be said to be reliable*.

Id. (emphasis added).

Validity, on the other hand, refers to the degree to which a measuring instrument measures what it purports to measure. *Id.*; *see also Horn*, 155 F.Supp.2d at 538-539. In short, reliability refers to consistency, while validity refers to accuracy or correctness. Thus, “forensic scientists or machines that are in agreement may be highly reliable (in agreement with each other) without being valid (without reaching the correct answer). *They can all be wrong.*” *Id.* This critical concept is one the government in this case has missed by equating testability with reproducibility rather than validity.

Toolmark identification, as the government concedes, is at its core a subjective endeavor. *See, e.g. United States v. Taylor*, 663 F.Supp.2d 1178 (D.N.M. 2009) (“Even the Government concedes that ‘the field continues to rely on a subjective match standard.’”); *United States v. Glynn*, 578 F.Supp.2d 567, 572 (S.D.N.Y. 2008)

("[B]allistics opinions are significantly subjective. Moreover, the standard defining when an examiner should declare a match - namely, "sufficient agreement" - is inherently vague."); NRC 2009 Report at 153-154 ("[T]he decision of the toolmark examiner remains a subjective decision based on unarticulated standards and no statistical foundation for estimation of error rates."); *Id.* at 155 ("A fundamental problem with toolmark and firearms analysis is the lack of a precisely defined process [for reaching identifications]."). The presence of high subjectivity coupled with vague standards makes testability a mirage. As explained below, the failure of the toolmarks field to objectively and empirically test the methodology in a scientific way exacerbates the problem, rendering opinion testimony in the field scientifically unreliable. Pursuant to *Daubert*, the testimony must be excluded.

B. PEER REVIEW AND PUBLICATION

The government suggests that peer review and publication of the theory of firearm and toolmark identification militates in favor of admission. Gov't Opp. at 8. Concededly, there is evidence to suggest that the AFTE theory of identification appears to be widely accepted by trained firearms examiners. But what Judge Saris learned is *Monteiro* is that the AFTE theory is not universally followed and is honored in the breach. According to Judge Saris:

At least in Massachusetts, however, the AFTE Theory does not appear to be a broadly recognized document. Apparently, while the AFTE Theory appears to be widely accepted by trained firearms examiners, it is not universally followed. Sgt. Weddleton testified that he had never before even seen or heard of it. Not only that, Mary Kate McGilvray, of the

Massachusetts State Police Crime Lab, also testified that she had never before read the AFTE Theory and that it was not the policy in her lab.

Monteiro, 407 F.Supp.2d at 370. In any event, the fact that there is a modicum of general acceptance within the toolmark community is not surprising, especially when one considers the community's incestuous affiliation with law enforcement and prosecutorial agencies. Forensic practitioners, like their academic counterparts, employ scientific techniques in order to find truth in particular contexts. However, their societal role differs considerably from academic scientists because their primary function is to afford a service to a customer by answering specific inquiries about evidence. In essence, forensic practitioners are service providers whose services are usually offered to law enforcement or prosecutorial agencies.¹ The role of "service provider" has significant repercussions, particularly concerning error-rate detection and general acceptance. Playing such a role has forced the forensic science community to develop a guild-like mentality. Forensic scientists frequently avoid answering questions regarding a technique's accuracy while also refraining from publicly criticizing a fellow colleague's questionable work. In short, a communal attitude has developed where practitioners are frowned upon by their peers when their publications reflect negatively on the forensic community's work product. Quelling any criticism achieves the community's ultimate goal of "promoting the

¹ The Scientific Working Group for Firearms and Toolmarks ("SWGGUN"), also cited by the government, suffers from the same incestuous relationship as AFTE. Twenty of the 22 members of SWGGUN's Board are law enforcement employees, and the board includes the President of AFTE. See http://www.swggun.org/swg/index.php?option=com_content&view=article&id=46&Itemid=3

impression that their techniques are accurate and reliable and that their conclusions are trustworthy.” William C. Thompson, *A Sociological Perspective on the Science of Forensic DNA Testing*, 30 U.C. Davis 1113, 1114 (1997). These credibility-enhancing labors, however, are antithetical to scientific self-scrutiny. *See Horn*, 155 F.Supp.2d at 556 (peer view “as contemplated by *Daubert* and *Kumho Tire* must involve critical analysis that can expose any weaknesses in the methodology or principles underlying the conclusions being reviewed”). More importantly, such a culture of professional self-protection severely inhibits forensic practitioners from identifying and rectifying problems associated not only with certain forensic methodologies but also with particular theories (e.g., the theory of individuality).

Given this background, the lack of critical analysis in peer-reviewed publications is not surprising. Because toolmark identification has no utility outside the courtroom it is unlikely to draw the interests of non-forensic scientists. For instance, DNA research is considered far superior to other forensic identification research because DNA has numerous applications outside the courtroom which draw the interests of neutral, non-law enforcement-related scientists. *See Green*, 405 F.Supp.2d at 109 n.7 (contrasting AFTE’s limited peer review among law enforcement employees with the neutral academics who participate in review of the DNA typing field). Second, in order for any forensic community to maintain its legitimacy and status in the courtroom, dissent must be kept to a bare minimum. As Professor Thompson explained:

The role of forensic scientists as service providers has important implications. Like most service providers, forensic scientists must convince

their clients that their services are valuable in order to succeed professionally. This imperative is most obvious for private laboratories, which often advertise and promote their services with commercial zeal. It is also important for employees of government laboratories who gain prestige and justify larger budgets by convincing public officials of the value of their services.

Because they need to sell their services, forensic scientists have incentives to put the best possible face on their work, to promote the impression that their techniques are accurate and reliable and that their conclusions are trustworthy. I believe that these incentives cause forensic scientists to act more like members of a trade guild than participants in a scientific discipline. For example, they may avoid openly raising questions about the reliability of forensic tests, avoid public discussion of technical problems or concerns, and refrain from publicly criticizing the work of other forensic scientists. They may also avoid publishing anything that might reflect negatively on their field, thereby making forensic science journals forums for self-promotion rather than self-criticism.

The problem with these credibility-enhancing efforts is that they are inconsistent with traditional forms of scientific self-scrutiny. This self-scrutiny is needed to maintain the quality of forensic science as a science. Forensic scientists may be slow to identify and solve problems with their techniques because problems are rarely discussed openly. The insularity of forensic science makes matters worse. The work of individual forensic scientists often receives little or no external scrutiny, therefore errors are often difficult to detect.

Thompson, *supra*, 1114-1115.

Perhaps the most obvious indicator that the toolmarks methodology lacks any scientific grounding flows from Judge Saris's observation that "[Daubert challenges] 'represent the biggest challenge facing the firearms discipline since it was firmly established in the 1920's.'" *Monterio*, 407 F.Supp.2d at 364 (*quoting* Sgt. Gerard Dutton, *Ethics in Forensic Firearms Investigation*, 37 Ass'n Firearm & Toolmark Examiners J. 79, 82 (2005)). If the field's scientific foundations were firmly established in the 1920s,

then why would *Daubert* challenges represent the greatest menace to the firearms community in more than eighty years? The truth, as authoritatively documented in the 2009 NAS Report, is that “many forensic tests — such as those used to infer the source of toolmarks...—have never been exposed to stringent scientific scrutiny,” that “the scientific knowledge base for toolmark and firearm analysis is fairly limited,” and that “[e]ven with more training and experience using new techniques, the decision of the toolmark examiner remains a subjective decision based on unarticulated standards and no statistical foundation for estimation of error rates.” NAS Report, p. 42, 153-154.

In sum, this Court cannot rely on peer review and publication as a basis for concluding that toolmark opinion evidence is reliable. As is the case with government’s suggestion of testability, *see supra*, the peer review process employed by firearms examiners is not the type of peer review envisioned by *Daubert* or *Kumho Tire*. The Court should therefore exclude the proposed opinion testimony and evidence.

C. ERROR RATES

When it comes to forensic identification evidence, perhaps the most important *Daubert* factor this Court must consider is whether the firearm community specifically, or the forensic science community generally, has conducted research to ascertain the accuracy (or inaccuracy) of its examiners. *See Daubert*, 509 U.S. at 594; *Kumho Tire*, 526 U.S. at 151. Here, the government writes that “the known or potential rate of error among trained firearm and toolmark examiners is quite low.” Gov’t Opp. at 9. What support there is for the government’s assertion, however, is anecdotal rather than the

result of any serious study. Neither the firearm community specifically nor the forensic science community generally has ever conducted double-blind independent proficiency testing aimed at ascertaining the accuracy (or inaccuracy) of firearm examiners. *See* Alfred Biasotti & John Murdoch, *The Scientific Basis of Firearms and Toolmark Identification*, in 3 *Modern Scientific Evidence* 215-216 (David. L. Faigman et al. eds. 1997). There is no explanation for this failure, which necessarily renders opinion testimony fundamentally unhelpful. As Judge Gertner explained:

The Court and ultimately the jury could still evaluate the testimony by considering the error rates in the field and the error rates for this examiner. In other words, even if his approach may be flawed, if examiners in the field manage to overcome those flaws, or if this examiner had a low error rate, the evidence may still be reliable, and the jury can evaluate it.

Green, 405 F.Supp.2d at 121.

Toolmark examiners try to rationalize the field's lack of research on error rates by naively asserting that "toolmark identification, because the process is so subjective and qualitative, it 'is not possible to calculate an absolute error rate for routine casework.'" *Monteiro*, 407 F.Supp.2d at 367 (quoting Richard Grzybowski et al., *Firearm/Toolmark Identification: Passing the Reliability Test Under Federal and State Evidentiary Standards*, 35 *Ass'n Firearm & Toolmark Examiners J.* 209, 219) (2003); Biasotti & Murdoch, *supra*, at 143 ("Based on the present data, the [toolmark] field is in a poor position to calculate error rates."). The "it's impossible to conduct error rate research" argument in effect permits toolmark and firearm examiners to use anecdotal evidence to demonstrate their alleged proficiency. Use of anecdotal evidence to bolster claims of

proficiency stifles error research. Under the current system, if an examiner generates an inclusionary or favorable finding for prosecutors, the prosecution team generally does not question the accuracy of this finding nor does it request additional testing to ensure the accuracy of the initial identification. *See Ramirez v. State*, 810 So.2d 836, 849 (Fla. 2001) (“Once a match is declared under his theory, no other knives are examined because an identification under this method purportedly eliminates all other knives in the world as possible sources of the wound.”). Rather, the prosecution simply assumes the correctness of the match and uses it to further bolster its case against the defendant. Even at trial, the accuracy of the examiner’s findings may not be exposed if defense counsel is not well-versed in forensic evidence or cross-examining forensic experts. Thus, an inaccuracy in an examiner’s identification may never surface.

The government concedes Judge Saris’s conclusion that it “is not possible to calculate an absolute error rate for routine casework” *Monteiro*, 407 F.Supp.2d at 367 (citation omitted). Gov’t Opp. at 9. *See also Taylor*, 663 F.Supp.2d at 1177 (“In his testimony at the *Daubert* hearing in this case, [the government’s expert] agreed that no actual error rate has been calculated for the field at this point.”); *United States v. Diaz*, 2007 WL 485967 at *9 (N.D.Cal. Feb 12, 2007) (“No true error rate will ever be calculated so long as the firearm-examiner community continues to rely on the subjective traditional pattern matching method of identification.”). Moreover, when proficiency testing is undertaken and the results are made public (which is rare), they suggest that novice and experienced examiners alike do not fully understand the scientific principles

and procedures they are attesting to in court, as many veteran forensic experts have failed routine, non-blind proficiency tests. *See* NAS 2009 Report at 44 (“As recently as September 2008, the Detroit Police crime laboratory was shut down following a Michigan State Police audit that found a 10 percent error rate in ballistic evidence.”). As the court in *Taylor*, noted: “Data from CTS [Collaborative Testing Services proficiency] testing done between 1978 and 1991 suggest that the rate of false identification is less than 1%. However, both Mr. Nichols and the Grzybowski article² acknowledge that uneven test administration, make-up, and level of difficulty significantly limit the usefulness of this result. Nonetheless, this number at least suggests that the error rate is quite low.” *Id.* at 1177. If “uneven test administration, make-up, and level of difficulty significantly limit the usefulness” of proficiency tests as measures of error rate, the government’s suggestion that the error rate is “quite low” cannot withstand serious scrutiny.

The firearm community’s reliance on anecdote and its concession that it has yet to and cannot conduct proficiency research weighs heavily in favor of non-admissibility for at least two reasons. First, because this type of error-rate evidence is non-existent the Government cannot definitively assert firearm examiners are highly accurate (or reliable) at linking a bullet to the one and only weapon which produced the marking. As a result, the Government cannot prove the reliability of firearm evidence and thus cannot satisfy

² The Grzybowski article cited by the *Taylor* court appears as Exhibit E2 to the Government’s Opposition, and warns that examiners “should be prepared to point out other problems with using the results of CTS proficiency tests to determine error rate.” Gov’t Opp., Exhibit E2 at 9.

Daubert. Second, as Judge Gertner wrote: “Expert evidence should not be excluded merely because witnesses practicing in that field make errors with some frequency, *but because the factfinder has no information about the likelihood of error in the opinions, and thus cannot adjust the weight to be given to the evidence.*” Green, 405 F.Supp.2d at 119 (emphasis added). Without “information about error rates, the initial factfinder, this Court, and the ultimate one, the jury, *have no accurate way of evaluating the testimony.*” *Id.* (emphasis added). If the court and jury are unable to accurately assess an expert’s opinion and properly weigh it, it goes without saying that the information offered by the expert is of no assistance because the jury has no way to judge what to do with this confusing (and highly prejudicial, see *infra*) information.

The government’s repetition of the toolmark field’s “quite low” error-rate mantra is unsupported and the lack of error rate renders the testimony in the field fundamentally flawed. Consequently, this Court should exclude the toolmark opinion testimony because it is neither reliable nor does it “assist the trier of fact to understand the evidence.” Fed. R. Evid. 702; *Daubert*, 509 U.S. at 591.

D. PATTERN MATCHING AND LACK OF STANDARDS

The government contends that toolmark testimony can be admitted because adequate standards control the process of making a pattern match and that method is widely accepted in the relevant community. Gov’t Opp. at 9-12. But no amount of “standards” change the fact that toolmark and firearms identifications are “based primarily on a visual inspection of patterns of toolmarks and is largely a subjective

determination based on experience and expertise.” *Monteiro*, 407 F.Supp.2d at 355. *See also United States v. Willock*, 696 F.Supp.2d 536, 560 (D.Md.2010) (“The AFTE’s most ardent supporter, Ronald Nichols of the ...[ATF] ... acknowledges the subjective component of toolmark examiners undertaking to discern ‘sufficient agreement’ in a toolmark identification, stating that ‘there is no universal agreement as to how much correspondence exceeds the best-known nonmatching situation.’”); *Taylor*, 663 F.Supp.2d at 1178 (“Even the Government concedes that ‘the field continues to rely on a subjective match standard.’”). Moreover, the likelihood of error in pattern matching is great because it is extremely difficult to “distinguish [] between class, subclass, and individual characteristics.” *Id.* *See also Monteiro*, 407 F.Supp.2d at 363 (“a firearm ‘may be wrongly identified as the source of a toolmark it did not produce if an examiner confuses subclass characteristics shared by more than one tool with individual characteristics unique to one and only one tool.’”) (citation omitted).

At the end of the day, “there are no national standards to be applied to evaluate how many marks must match.” *Green*, 405 F.Supp.2d at 114. *See also Taylor*, 663 F.Supp.2d at 1177 (“The AFTE Theory, thus, does not provide any uniform numerical standard examiners can use to determine whether or not there is a match and, indeed, Mr. Nichols indicated in his testimony that most AFT examiners do not use any numerical standard. Instead, the AFTE theory is circular.”). As Judge Saris commented, this definition, for all intents and purposes, represents a tautology:

Instead, the AFTE Theory, upon which the government relies, is tautological: it requires each examiner to decide when there is 'sufficient agreement' of toolmarks to constitute an 'identification.' This threshold is surpassed when the examiner finds that the agreement of toolmarks 'exceeds the best agreement demonstrated between toolmarks known to have been produced by different tools and is consistent with agreement demonstrated by toolmarks known to have been produced by the same tool.' Toolmark analysis does not follow an objective standard requiring, say, a certain percentage of marks to match. Rather, as noted, this 'threshold is currently held in the mind's eye of the examiner and is based largely on training and experience.'"

Monteiro, 407 F.Supp.2d at 370. See *Glynn*, 578 F.Supp.2d at 574 ("[B]allistics comparison lacks defining standards to a degree that exceeds most other kinds of forensic expertise. For example, whereas both a ballistics examiner and a fingerprint examiner are ultimately called upon to make a subjective judgment of whether the agreement between two pieces of evidence is "sufficient" to constitute a "match," a fingerprint examiner may not declare a match unless a pre-specified number of "points" of similarity exist between the two samples....Although attempts been made to introduce similar minimum standards and 'protocols' into ballistics analysis, such attempts have not yet met with general acceptance...")

It was precisely this kind of unguided pattern matching evidence that was criticized by the NAS 2009 Report. It found that the "the decision of the toolmark examiner remains a subjective decision based on unarticulated standards and no statistical foundation for estimation of error rates."), *id* at 153-154, leading to the conclusion that the "fundamental problem with toolmark and firearms analysis is the lack of a precisely defined process [for reaching identifications]." *Id.* at 155. While the

pattern matching methods may be widely accepted in the toolmarks community, the “standards” do not insure any measure of scientific reliability. Where the testimony is based on such unreliable foundations, the opinion evidence must be excluded.

CONCLUSION

The government is indeed correct that no court has yet completely excluded toolmark identification evidence from trial. Gov’t Opp. at 5. As argued in the defendant’s initial motion, however, the changes engendered by *Daubert* and the amendment of Rule 702 require reexamination of the long-held practice of permitting testimony in the scientifically suspect toolmarks examination field. In a capital case, where the need for evidentiary reliability is at its peak, the Court must be exceptionally careful when considering whether to admit or exclude evidence. *United States v. Taveras*, 424 F. Supp. 2d 446, 462 (E.D.N.Y. 2006). Testability, peer review, error rate, standards, subjectivity and general acceptance as to toolmarks pattern matching are all, to varying degrees, problematic and the Court should exclude the testimony. At a minimum, it should hold an evidentiary hearing to resolve this motion.

Respectfully submitted,

DZHOKHAR TSARNAEV
By his attorneys



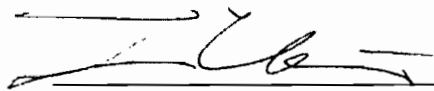
Judy Clarke, Esq. (CA Bar # 76071)
CLARKE & RICE, APC
1010 Second Avenue, Suite 1800
San Diego, CA 92101
(619) 308-8484
JUDYCLARKE@JCSRLAW.NET

David I. Bruck, Esq.
220 Sydney Lewis Hall
Lexington, VA 24450
(540) 460-8188
BRUCKD@WLU.EDU

Miriam Conrad, Esq. (BBO # 550223)
Timothy Watkins, Esq. (BBO # 567992)
William Fick, Esq. (BBO # 650562)
FEDERAL PUBLIC DEFENDER OFFICE
51 Sleeper Street, 5th Floor
(617) 223-8061
MIRJAM_CONRAD@FD.ORG
TIMOTHY_WATKINS@FD.ORG
WILLIAM_FICK@FD.ORG

Certificate of Service

I hereby certify that this document was served on counsel for the government by email and paper copy by interoffice delivery on December 30, 2014.



Timothy Watkins